John von Neumann Institute for Computing

Semiempirical Methods

Walter Thiel

published in

Modern Methods and Algorithms of Quantum Chemistry,
Proceedings, Second Edition, J. Grotendorst (Ed.),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **3**, ISBN 3-00-005834-6, pp. 261-283, 2000.

http://www.fz-juelich.de/nic-series/

SEMIEMPIRICAL METHODS

WALTER THIEL
Max-Planck-Institut für Kohlenforschung
Kaiser-Wilhelm-Platz 1
45470 Mülheim
Germany
E-mail: thiel@mpi-muelheim.mpg.de

Recent methodological advances in semiempirical quantum chemistry and associated algorithmic developments are reviewed. After a brief overview over the current status of established semiempirical methods, the following topics are covered: new general-purpose methods with orthogonalization corrections, MNDO–NMR chemical shifts, analytic derivatives, parallelization of semiempirical codes, and linear scaling vs combined QM/MM approaches.

1 Introduction

This article outlines a lecture held at the winterschool on "Modern Methods and Algorithms of Quantum Chemistry" at Jülich (February 2000). It summarizes some recent developments in the field of semiempirical quantum chemistry, with emphasis on the work from our group. Theoretical derivations and detailed numerical results are generally not presented since they can be found in the original publications.

Over the past decades the semiempirical molecular orbital (MO) methods have been used widely in computational studies. There are several books [1,2,3,4,5,6] and reviews [7,8,9,10,11,12,13,14] which describe the underlying theory, the different variants of semiempirical methods, and the numerical results. Semiempirical approaches are normally formulated within the same conceptual framework as ab initio methods, but they neglect many smaller integrals to speed up the calculations. In order to compensate for the errors caused by these approximations, empirical parameters are introduced into the remaining integrals and calibrated against reliable experimental or theoretical reference data. This strategy can only be successful if the semiempirical model retains the essential physics to describe the properties of interest. Provided that this is the case, the parameterization can account for all other effects in an average sense, and it is then a matter of validation to establish the numerical accuracy of a given approach.

In current practice, semiempirical methods serve as efficient computational tools which can yield fast quantitative estimates for a number of properties. This may be particularly useful for correlating large sets of experimental and theoretical data, for establishing trends in classes of related molecules, and for scanning a computational problem before proceeding with higher-level treatments. Compared with ab initio or density functional methods, semiempirical calculations are much faster, typically by several orders of magnitude[14], but they are also less accurate, with errors that are less systematic and thus harder to correct. Hence, there remains the need to improve semiempirical methods with regard to their accuracy and range of applicability, without compromising their computational efficiency. In addition, there is the need to develop new algorithms in order to exploit modern computer architectures and

to extend semiempirical calculations to ever larger molecules.

2 Established methods

Quantum-chemical semiempirical treatments are defined by the following specifications:

(a) *The underlying theoretical approach:* Most current general-purpose semiempirical methods are based on MO theory and employ a minimal basis set for the valence electrons. Electron correlation is treated explicitly only if this is necessary for an appropriate zero-order description.

(b) *The integral approximation and the types of interactions included:* Traditionally there are three levels of integral approximation [2,15] – CNDO (complete neglect of differential overlap), INDO (intermediate neglect of differential overlap), and NDDO (neglect of diatomic differential overlap). NDDO is the best of these approximations since it retains the higher multipoles of charge distributions in the two-center interactions (unlike CNDO and INDO which truncate after the monopole).

(c) *The integral evaluation:* At a given level of integral approximation, the integrals are either determined directly from experimental data or calculated from the corresponding analytical formulas or computed from suitable parametric expressions. The first option is generally only feasible for the one-center integrals which may be derived from atomic spectroscopic data. The choice between the second and third option is influenced by the ease of implementation of the analytical formulas, but mainly depends on an assessment of how to model the essential interactions.

(d) *The parameterization:* Semiempirical MO methods are parameterized to reproduce experimental reference data (or, possibly, accurate high-level theoretical predictions as substitutes for experimental data). The reference properties are best selected such that they are representative for the intended applications. The quality of semiempirical results is strongly influenced by the effort put into the parameterization.

In our terminology, the specifications (a)–(b) define a semiempirical model, (a)–(c) an implementation of a given model, and (a)–(d) a particular method.

Over the years, a large number of methods with different choices for (a)–(d) and different acronyms have been published, including CNDO/2 [16], CNDO/S [17], INDO [18], MINDO/3 [19], INDO/S [20,21], SINDO1 [22,23], MSINDO [24,25], MNDO [26,27], MNDOC [28], AM1 [29], PM3 [30], SAM1 [31,32], MNDO/d [33,34], PM3/tm [35], and NDDO–G [36]. The most popular semiempirical methods for studying ground-state potential surfaces are based on the MNDO model [26]. As a point of reference for the further discussion, we therefore first outline the MNDO formalism for closed-shell molecules.

MNDO is a valence-electron self-consistent-field (SCF) MO treatment which employs a minimal basis of atomic orbitals (AOs, ϕ_μ) and the NDDO integral

approximation. The molecular orbitals ψ_i and the corresponding orbital energies ε_i are obtained from the solution of the secular equations ($S_{\mu\nu} = \delta_{\mu\nu}$ for NDDO):

$$\psi_i = \sum_{\mu} c_{\mu i} \phi_{\mu} \quad , \tag{1}$$

$$0 = \sum_{\nu} \left(F_{\mu\nu} - \delta_{\mu\nu} \varepsilon_i \right) c_{\nu i} \quad . \tag{2}$$

Using superscripts to assign an AO (with index μ, ν, λ, σ) to an atom A or B, the NDDO Fock matrix elements $F_{\mu\nu}$ are given as

$$F_{\mu^A \nu^A} = H_{\mu^A \nu^A} + \sum_{\lambda^A} \sum_{\sigma^A} P_{\lambda^A \sigma^A} \left[\left(\mu^A \nu^A, \lambda^A \sigma^A \right) - \frac{1}{2} \left(\mu^A \lambda^A, \nu^A \sigma^A \right) \right] + \sum_{B} \sum_{\lambda^B} \sum_{\sigma^B} P_{\lambda^B \sigma^B} \left(\mu^A \nu^A, \lambda^B \sigma^B \right) \quad , \tag{3}$$

$$F_{\mu^A \nu^B} = H_{\mu^A \nu^B} - \frac{1}{2} \sum_{\lambda^A} \sum_{\sigma^B} P_{\lambda^A \sigma^B} \left(\mu^A \lambda^A, \nu^B \sigma^B \right) \quad , \tag{4}$$

where $H_{\mu\nu}$ and $P_{\lambda\sigma}$ are elements of the one-electron core Hamiltonian and the density matrix, respectively, and $(\mu\nu, \lambda\sigma)$ denotes a two-electron integral. The total energy E_{tot} of a molecule is the sum of its electronic energy E_{el} and the repulsions E_{AB}^{core} between the cores of all atoms A and B.

$$E_{el} = \frac{1}{2} \sum_{\mu} \sum_{\nu} P_{\mu\nu} \left(H_{\mu\nu} + F_{\mu\nu} \right) \quad , \tag{5}$$

$$E_{tot} = E_{el} + \sum_{A\langle} \sum_{B} E_{AB}^{core} \quad . \tag{6}$$

It is obvious from Eqs. (1)–(6) that the MNDO model includes only one-center and two-center terms which accounts for much of its computational efficiency. Current implementations of the MNDO model (e.g., in the MNDO, AM1, and PM3 methods) are quite similar: Conceptually the one-center terms are taken from atomic spectroscopic data, with the refinement that slight adjustments are allowed in the optimization to account for possible differences between free atoms and atoms in a molecule. The one-center two-electron integrals derived from atomic spectroscopic data are considerably smaller than their analytically calculated values which is (at least partly) attributed to an average incorporation of electron correlation effects. For reasons of internal consistency, these integrals provide the one-center limit ($R_{AB} = 0$) of the two-center two-electron integrals $\left(\mu^A \nu^A, \lambda^B \sigma^B \right)$, whereas the asymptotic limit of $\left(\mu^A \nu^A, \lambda^B \sigma^B \right)$ for $R_{AB} \rightarrow \infty$ is determined by classical electrostatics. The semiempirical calculation of $\left(\mu^A \nu^A, \lambda^B \sigma^B \right)$ conforms to these limits and evaluates these integrals from semiempirical multipole-multipole interactions [33,37]: The relevant multipoles are represented by suitable point-charge configurations whose interaction is damped according to the Klopman-Ohno formula. Therefore, at intermediate distances, the semiempirical two-electron integrals are smaller than their analytical counterparts which again reflects some inclusion of

Table 1. Mean Absolute Errors for Organic Molecules (C, H, N, O).

Property[a]	N[b]	MNDO	AM1	PM3
ΔH_f (kcal/mol)	133	6.3	5.5	4.2
R(Å)	228	0.015	0.017	0.011
θ (deg)	92	2.69	2.01	2.22
IP (eV)	51	0.47	0.36	0.43
μ (D)	57	0.32	0.25	0.27

a Heats of formation ΔH_f, bond lengths R, bond angles θ, ionization potentials IP (Koopmans' theorem), dipole moments μ.
b Number of comparisons.

electron correlation effects. Aiming for a reasonable balance between electrostatic attractions and repulsions within a molecule, the core-electron attractions and the core-core repulsions are treated in terms of the corresponding two-electron integrals, neglecting, for example, penetration effects. The additional effective atom-pair potential that is included in the core-core repulsions (with an essentially exponential repulsion in MNDO and a more flexible parametric function in AM1 and PM3) attempts to compensate for errors introduced by the above assumptions, but mainly represents the Pauli exchange repulsions. Covalent bonding arises from the two-center one-electron integrals $H_{\mu^A \nu^B}$ (resonance integrals) which are often taken to be proportional to the corresponding overlap integrals.

The MNDO model in its current standard implementations (MNDO, AM1, PM3) has been parameterized primarily with respect to ground-state properties, with particular emphasis on the energies and geometries of organic molecules. It has become common practice to judge the accuracy of computational methods from statistical evaluations for standard validation sets. Such results are shown in Table 1 for organic compounds using the original MNDO set [27], in Table 2 for second-row and heavier molecules using the MNDO/d set [34], and in Table 3 for first-row and second-row compounds using the recent G2 neutral test set that is mainly employed in ab initio and density functional (DFT) work [38]. The data in Tables 1–3 have been assembled for previous reviews [14,39].

The results in Table 1 indicate that AM1 and PM3 offer some improvement in accuracy over the original MNDO method, but the mean absolute errors remain of the same order of magnitude. Hence, AM1 and PM3 may be viewed as variants of MNDO that explore the limits of the underlying theoretical model through careful parameterization.

In general, the errors in semiempirical calculations for organic compounds (Table 1) are smaller than those for molecules containing second-row and heavier elements (Table 2). As expected, the methods with an sp basis (MNDO, AM1, PM3) fail for hypervalent compounds, which are described more accurately upon extension of the MNDO model to d orbitals (MNDO/d, see Table 2).

Concerning the G2 validation set, it is obvious from Table 3 that the G2 approach is the most accurate one among those studied (as anticipated), followed by G2(MP2) and B3LYP. The semiempirical methods (especially PM3 and MNDO/d)

Table 2. Mean Absolute Errors for Molecules Containing Second-Row and Heavier Elements[a]

Property[b]	N[c]	MNDO	AM1	PM3	MNDO/d
ΔH_f (kcal/mol)	575	29.2	15.3	10.9	5.4
R(Å)	441	0.072	0.063	0.065	0.056
θ (deg)	243	3.7	3.4	7.4	2.5
IP (eV)	200	0.89	0.55	0.64	0.45
μ (D)	133	0.55	0.50	0.60	0.35
Hypervalent compounds only					
ΔH_f (kcal/mol)	67	143.2	61.3	19.9	5.4

a Elements: Na, Mg, Al, Si, P, S, Cl, Br, I, Zn, Cd, and Hg are included. Original data see ref. [34]

b Heats of formation ΔH_f, bond lengths R, bond angles θ, ionization potentials IP (Koopmans' theorem), dipole moments μ.

c Number of comparisons for MNDO/d (slightly lower for the other methods due to missing parameters, see ref. [34] for details).

Table 3. Mean Absolute Errors Δ_{abs}, Largest Positive Errors Δ_{pos}, and Largest Negative Errors Δ_{neg} of Heats of Formation (kcal/mol) for the Molecules from the G2 Neutral Test Set[a].

Method	Δ_{abs}	Δ_{pos}	Δ_{neg}	Ref.
G2	1.58	8.2	-7.1	[38]
G2(MP2)	2.04	10.1	-5.3	[38]
LDA(SVWN)	91.16	228.7	none	[38]
BLYP	7.09	28.4	-24.8	[38]
BP86	20.19	49.7	-6.3	[38]
B3LYP	3.11	8.2	-20.1	[38]
MNDO	9.32	27.6	-116.7	[39]
AM1	7.81	42.5	-58.2	[39]
PM3	7.01	23.1	-32.2	[39]
MNDO/d	7.26	27.6	-33.9	[39]

a The G2 neutral test set contains 148 molecules [38]. Due to the lack of parameters for certain elements, the data for MNDO, AM1, and PM3 refer to 146, 142, and 144 molecules, respectively.

show similar errors as BLYP, whereas BP86 and particularly LDA(SVWN) overbind strongly. To put these results into perspective, it should be noted that the complete geometry optimization of all 148 test molecules took altogether less than 20 seconds on an SGI R10000 workstation, for any of the four semiempirical methods. Since this computational effort is several orders of magnitude lower than that at the ab initio or DFT levels, the overall performance of the semiempirical methods appears acceptable (Table 3).

The statistical evaluations in Tables 1–3 (and many others that are available in the literature [11,12,13,30,34]) indicate that the established semiempirical methods can often be applied with useful accuracy and at very low computational costs. The

following limitations should be kept in mind, however:

(a) In general, errors tend to be more systematic at a given ab initio or DFT level and may therefore often be taken into account by suitable corrections. Errors in semiempirical calculations are normally less uniform and thus harder to correct.

(b) The accuracy of the semiempirical results may be different for different classes of compounds, and there are elements that are more "difficult" than others. Such variations in the accuracy are again less pronounced in high-level ab initio and DFT calculations.

(c) Semiempirical methods can only be applied to molecules containing elements that have been parameterized, while ab initio and DFT methods are generally applicable (apart from technical considerations such as basis set availability).

(d) Semiempirical parameterizations require reliable experimental or theoretical reference data and are impeded by the lack of such data. Such problems do not occur in ab initio or DFT approaches.

(e) Different parameterizations of a given semiempirical model may be required for different properties to obtain useful accuracy, and there is no systematic procedure for improving the results, unlike ab initio methods which provide a convergent path to the exact solution of the nonrelativistic Schrödinger equation.

To overcome some of these limitations, it seems desirable to develop better semiempirical methods. In the next sections, two such attempts are presented: First, a general-purpose parameterization of an approach that goes beyond the MNDO model by explicitly including orthogonalization effects, and secondly, a special parameterization of the MNDO model for the calculation of NMR chemical shifts.

3 Beyond the MNDO model: Orthogonalization corrections

As discussed above, the established MNDO-type methods do not treat the Pauli exchange repulsions explicitly, but attempt to incorporate them through an effective atom-pair potential that is added to the core-core repulsion. When trying to improve the MNDO model, it would seem logical to include the Pauli exchange repulsions explicitly in the electronic calculation and to remove the effective atom-pair potential from the core-core repulsion. For the sake of consistency, other one-electron terms of similar magnitude should then also be treated explicitly, i.e. penetration integrals and core-valence interactions (effective core potentials).

The zero-differential-overlap (ZDO) approximation causes the formal neglect of the Pauli exchange repulsions in semiempirical methods. Its consequences have been analyzed in many studies (see e.g. refs. [40,41,42,43,44,45,46,47]). ZDO-related deficiencies include:

(a) The gaps between bonding and antibonding molecular orbitals, as well as the corresponding excitation energies, are significantly underestimated.

(b) The correct pairing properties for the electronic states of of conjugated hydrocarbons are not reproduced [45].

(c) Barriers to internal rotation as in ethane are underestimated, with implications for other conformational properties (for a detailed discussion see ref. [46]).

(d) The closed-shell repulsions due to four-electron two-orbital interactions are not recovered properly, which causes problems, for example, with antiaromatic systems, intermolecular interaction potentials, and certain transition structures [41,42,43,44,45,46,47].

To account for the Pauli exchange repulsions more explicitly at the semiempirical level, orthogonalization corrections need to be addressed. The chosen strategy is based on the following general considerations [47].

Ab initio SCF methods solve the Roothaan-Hall pseudo-eigenvalue problem

$$\mathbf{F}\ \mathbf{C} = \mathbf{S}\ \mathbf{C}\ \mathbf{E} \quad , \tag{7}$$

where **F**, **C**, and **S** denote the Fock, eigenvector and overlap matrix, respectively, and **E** is the diagonal matrix of orbital energies. Orthogonalization of the basis leads to a standard eigenvalue problem

$${}^{\lambda}\mathbf{F}{}^{\lambda}\mathbf{C} = {}^{\lambda}\mathbf{C}\mathbf{E} \quad , \tag{8}$$

where the superscript λ denotes a quantity expressed in an orthogonalized basis. The corresponding transformation can be achieved through a symmetric Löwdin orthogonalization. By contrast, semiempirical methods solve a secular equation,

$${}^{NDDO}\mathbf{F}\ {}^{NDDO}\mathbf{C} = {}^{NDDO}\mathbf{C}\ \mathbf{E} \quad , \tag{9}$$

where the transformation **F** to ${}^{\lambda}\mathbf{F}$ is not explicitly performed. This suggests that the semiempirical Fock matrix implicitly refers to an orthogonal basis:

$${}^{NDDO}\mathbf{F} \approx {}^{\lambda}\mathbf{F} \tag{10}$$

The neglect of all three-center and four-center two-electron integrals in NDDO approximation [2,15] is consistent with this interpretation because these integrals are vanishingly small only in an orthogonalized basis. The Fock matrix contains both one-electron (**H**) and two-electron (**G**) terms, which can be handled separately during Löwdin orthogonalization:

$${}^{\lambda}\mathbf{F} = {}^{\lambda}\mathbf{H} + {}^{\lambda}\mathbf{G} \tag{11}$$

$${}^{\lambda}\mathbf{H} = \mathbf{S}^{-1/2}\mathbf{H}\mathbf{S}^{-1/2} \tag{12}$$

$${}^{\lambda}\mathbf{G} = \mathbf{S}^{-1/2}\mathbf{G}\mathbf{S}^{-1/2}. \tag{13}$$

To account for the orthogonalization effects arising from these transformations, different strategies may be followed in a semiempirical context. The direct use of Eq. (13) is not feasible since it requires the prior calculation of all two-electron integrals and would thus sacrifice the computational efficiency of the NDDO approach. On the other hand, the exact orthogonalization of the one-electron part according to Eq. (12) would be feasible computationally. However, using Eq. (12) without Eq. (13) introduces an imbalance between the one-electron and two-electron parts of the

Fock matrix and is therefore problematic. In the literature, several variants of this approach have been studied, where different parts of the Fock matrix are subjected to the exact Löwdin transformation while the ZDO approximation is applied to the remainder (see e.g. refs. [45,48,49]). To our knowledge, none of these attempts has been successfully incorporated into a general-purpose semiempirical method. Therefore, we have adopted the alternative strategy of representing the dominant one-electron orthogonalization corrections by suitable parametric functions. These corrections can then be adjusted during the parameterization process, as in previous approaches at the CNDO and INDO levels [22,23,24,25,50].

These basic ideas have been implemented in two steps. First, the Pauli exchange repulsions have been introduced as valence-shell orthogonalization corrections only in the one-center part of the core Hamiltonian [51,52]. In the second step, they have also been incorporated in the two-center part of the core Hamiltonian [46,47], i.e. in the resonance integrals. Both developments have been guided by analytic ab initio formulas and numerical ab initio SCF results. Their implementations are actually quite similar: In both cases, a Gaussian minimal basis set is used for technical reasons, and most two-center interactions are evaluated analytically followed by an appropriate Klopman-Ohno scaling. The valence-shell orthogonalization corrections are represented in terms of the resonance integrals through a truncated and parameterized series expansion. The resonance integrals contain a parameterized radial part while the angular part is the same as in the corresponding overlap integral.

The first approach [51,52] which we now call OM1 (orthogonalization method 1) contains only one-center and two-center terms, since the dominant orthogonalization corrections to the one-center part of the core Hamiltonian involve only a second atom. By contrast, the second approach [46,47] labelled OM2 (orthogonalization method 2) includes three-center contributions, since the corrections to the resonance integrals $\beta_{\mu\lambda}$ involve a third atom. The relevant corrections in OM2 are given by [46,47]

$$^{\lambda}H_{\mu\lambda} = \beta_{\mu\lambda} - \frac{1}{2}\sum_{\rho}^{C}(S_{\mu\rho}\beta_{\rho\lambda} + \beta_{\mu\rho}S_{\rho\lambda}) + \frac{1}{8}\sum_{\rho}^{C} S_{\mu\rho}S_{\rho\lambda}(H_{\mu\mu} + H_{\lambda\lambda} - 2H_{\rho\rho}) \ . \quad (14)$$

These three-center contributions reflect the stereochemical environment of each electron pair bond and should thus be important for modeling conformational properties. Technically, their inclusion increases the computational effort, but does not affect its scaling behavior since the three-center terms fall off like the square of an overlap integral and can therefore safely be neglected beyond certain cutoffs.

OM1 has been parameterized for the elements H, C, N, O, and F [51,52]. Significant qualitative improvements over the established MNDO-type methods are found in several areas, particularly for excited states. The mean absolute error in vertical excitation energies is 0.28 eV, much lower than in AM1 (1.20 eV) or PM3 (1.18 eV). This is readily rationalized: The destabilization of antibonding molecular orbitals is greater than the stabilization of bonding molecular orbitals at the ab initio level. This effect is not taken into account in the established MNDO-type methods whereas it is incorporated in OM1 through the orthogonalization correc-

Table 4. Mean Absolute Errors of Heats of Formation (kcal/mol) for First-Row Molecules from the G2 Neutral Test Set[a].

Method	First-row	CH	CHN	CHNO	CHNOF
G2	1.53(93)	1.31(30)	1.04(17)	1.37(33)	3.29(12)
BLYP	7.38	7.16	5.41	8.34	8.70
B3LYP	2.42	2.66	2.23	2.18	2.86
MNDO	7.71	7.42	5.93	9.13	7.62
AM1	7.44	6.57	5.05	9.35	7.89
PM3	6.86	6.43	6.75	6.88	7.46
OM1	4.64	3.20	3.67	5.89	6.37
OM2	3.36(81)	2.37	3.25	4.30	[b]

a The number of molecules in a subgroup is given in parentheses. Results for G2, BLYP, and B3LYP have been derived from the published data [38].
b OM2 parameters for F still missing.

tions [14]. Hence, the excitation energies are raised in a natural manner by correcting for deficiencies inherent to the ZDO approximation.

OM2 has presently been parameterized for H, C, N, and O [46]. As expected theoretically (see above), the numerical results show qualitative improvements for conformational properties, including rotational barriers, relative energies of isomers, and ring conformations. Hydrogen bonds are generally described much better than previously. The barriers for typical pericyclic reactions are realistic: for example, in the Diels-Alder reaction between butadiene and ethylene, the barrier is lower for the concerted pathway than for the biradicaloid one, and it decreases with increasing cyano substitution (in agreement with experiment and ab initio results).

While these qualitative advances are gratifying, it is also important to demonstrate that OM1 and OM2 perform well for the usual ground-state properties (see above). This is indeed the case [39,46,52]. OM1 and OM2 offer consistent small improvements over the established MNDO-type models, as can be seen, for example, from Table 4 which presents a statistical evaluation of thermochemical results for the first-row compounds and suitable subgroups from the G2 neutral test set [38,39].

Judging from the presently available OM1 and OM2 results [38,46,52,53], the explicit inclusion of Pauli exchange repulsions has led to qualitative and quantitative improvements in several important areas, which can partly be traced back to improvements in the underlying theoretical model. More work is needed to explore the limitations of these new approaches and to extend them to heavier elements.

4 NMR chemical shifts

NMR chemical shifts are an important source of information on molecular structure and reactivity in many fields of chemistry. They can be measured with high accuracy and sensitivity for increasingly complex systems, including large biomolecules whose solution structures can be determined from NMR chemical shifts and coupling constants. The accurate theoretical calculation of absolute NMR shieldings

is a challenging task which is best addressed by high-level theoretical methods. A large number of ab initio and density functional approaches to the evaluation of the NMR chemical shifts have indeed become available in recent years. Most of these approaches are based on the GIAO (Gauge-Including Atomic Orbitals) [54] and IGLO (Individual Gauge for Localized Orbitals) [55,56] concepts. Many such ab initio and DFT calculations have been published, and several reviews of such work have appeared [56,57,58].

Despite these tremendous advances at the ab initio and DFT level, a reliable and less costly semiempirical method would clearly be useful for the treatment of larger systems and the rapid estimation of chemical shifts in cases where the full versatility of ab initio and DFT methods is not needed. Even though there has been some previous semiempirical work in this area (for a review of the earlier literature, see refs. [59,60]), a reliable semiempirical method for computing NMR chemical shifts appears to be lacking. We have therefore decided to implement the evaluation of the NMR chemical shift tensor at the MNDO level for an spd basis set using gauge-including atomic orbitals and analytic derivative theory, followed by a specific parameterization of the GIAO–MNDO model with respect to experimental reference data [59,60]. In the following, we outline this development and summarize the results that have been obtained so far (for details see ref. [60]).

The NMR shielding tensor components σ_{ab} are second-order properties given by the mixed derivative of the energy with respect to the components of the nuclear magnetic moment $\vec{\mu}$ and the magnetic field strength $\vec{B}$:

$$\sigma_{ab} = \frac{\partial^2 E}{\partial B_a \partial \mu_b}\Big|_{\vec{\mu},\vec{B}=0} \tag{15}$$

For the evaluation of the shielding tensor it is sufficient to include only the terms linear and bi-linear in $\vec{\mu}$ and $\vec{B}$ in the one-electron Hamiltonian, which is then given by (atomic units) [54,61]:

$$\hat{h} = \hat{h}^{00} + i\sum_a \hat{h}^{a0} B_a + i\sum_b \hat{h}^{0b}\mu_b + \sum_{a,b} \hat{h}^{ab} B_a \mu_b \tag{16}$$

In the GIAO approach, the origin independence of the chemical shift is ensured by introducing a field-dependent factor into the basis functions [62]:

$$\varphi\left(\vec{r}-\vec{R}\right) = \chi\left(\vec{r}-\vec{R}\right)\exp\left(-\frac{i}{2c}\left(\vec{B}\times\vec{R}\right)\cdot\vec{r}\right) \tag{17}$$

where $\chi\left(\vec{r}-\vec{R}\right)$ is a field-independent basis function centered at $\vec{R}$.

GIAO basis functions do not depend on the nuclear magnetic moment, so that differentiation with respect to $\vec{\mu}$ may be performed using the Hellmann-Feynman theorem. Subsequent differentiation with respect to $\vec{B}$ introduces terms due to the derivatives of the Hamiltonian, the GIAOs, and the wave function. In the density matrix formulation, the resulting expression is given by [61]:

$$\sigma_{ab} = \sum_{\mu\nu} P_{\mu\nu} H^{ab}_{\mu\nu} - \sum_{\mu\nu} P^a_{\mu\nu} H^{0b}_{\mu\nu} \tag{18}$$

The purely real matrix elements $H^{ab}_{\mu\nu}$ and $H^{0b}_{\mu\nu}$ are:

$$H^{ab}_{\mu\nu} = \frac{\partial}{\partial B_a} \left\langle \varphi_\mu \middle| \frac{\partial \hat{h}}{\partial \mu_b} \middle| \varphi_\nu \right\rangle \Bigg|_{\vec{\mu},\vec{B}=0} \tag{19}$$

$$H^{0b}_{\mu\nu} = \left\langle \chi_\mu \middle| \frac{\partial \hat{h}}{\partial i\mu_b} \middle| \chi_\nu \right\rangle \Bigg|_{\vec{\mu},\vec{B}=0} \tag{20}$$

The imaginary part of the first-order density matrix $\mathbf{P}^a$ is formally defined as derivative of the zero-order density matrix $\mathbf{P}$:

$$P^a_{\mu\nu} = \frac{\partial P_{\mu\nu}}{\partial i B_a} \tag{21}$$

The perturbed density matrix is obtained by solving the CPHF (coupled perturbed Hartree-Fock) equations, which take a particularly simple form in the MNDO approximation [59,61]:

$$\mathbf{F}^a\mathbf{P} + \mathbf{F}\mathbf{P}^a = \mathbf{P}^a\mathbf{F} + \mathbf{P}\mathbf{F}^a \tag{22}$$

$$\mathbf{P}^a\mathbf{P} + \mathbf{P}\mathbf{P}^a = 2\mathbf{P}^a \tag{23}$$

$$\mathbf{F}^a = \mathbf{H}^a + \mathbf{G}\left(\mathbf{P}^a\right) \tag{24}$$

$\mathbf{H}^a$ is the imaginary part of the static derivative of the one-electron Hamiltonian with respect to the magnetic field:

$$\begin{aligned} H^a_{\mu\nu} &= \frac{\partial}{\partial i B_a} \left\langle \varphi_\mu \middle| \hat{h} \middle| \varphi_\nu \right\rangle \Big|_{\vec{\mu},\vec{B}=0} \\ &= \left\langle \chi_\mu \middle| \hat{h}^{a0} - \frac{1}{2c} \left(\vec{r} \times \vec{R_\mu}\right)_a \hat{h}^{00} + \frac{1}{2c} \hat{h}^{00} \left(\vec{r} \times \vec{R_\nu}\right)_a \middle| \chi_\nu \right\rangle \end{aligned} \tag{25}$$

$\mathbf{G}\left(\mathbf{P}^a\right)$ is the analogue of the two-electron part of the Fock matrix built using the first-order density matrix in place of the unperturbed density. Equation (24) is considerably simpler than its ab initio analogue [61] due to the cancellation of the field-dependent factors in all two-electron integrals that are retained in MNDO approximation. The CPHF equations (22)–(24) can be solved directly in the AO (atomic orbital) basis or transformed into the explicitly linear non-redundant form [59]. In either case, the equations are solved by a rapidly converging iterative procedure.

In ab initio approaches, the matrix elements $H^a_{\mu\nu}$ in Eq. (25) are evaluated analytically. MNDO-type methods assume the two-center matrix elements $H_{\mu\nu}$ of the operator $\hat{h}^{00}$ to be proportional to the corresponding overlap integral $S_{\mu\nu}$:

$$H_{\mu\nu} = \left\langle \chi_\mu | \hat{h}^{00} | \chi_\nu \right\rangle \approx b_{\mu\nu} S_{\mu\nu} = H^{\mathrm{MNDO}}_{\mu\nu} \tag{26}$$

Given this choice, a consistent approximation is required in Eq. (25) in order to preserve the origin independence of the results. For reasons discussed in more detail

elsewhere [59,60], we have adopted the following semiempirical expression for $H^a_{\mu\nu}$:

$$H^a_{\mu\nu}(\mathrm{MNDO}) = \frac{1}{2c}\Bigg\{ \left(\vec{R}_\mu \times \vec{R}_\nu\right) H^{\mathrm{MNDO}}_{\mu\nu} + \left(\vec{R}_\mu - \vec{R}_\nu\right) \times \left(b_{\mu\nu} \left\langle \chi_\mu \left| \vec{r} - \vec{R}_\nu \right| \chi_\nu \right\rangle \right) - \frac{1}{2} \left\langle \chi_\mu \left| \hat{L}^{R_\nu} \right| \chi_\nu \right\rangle + \frac{1}{2} \left\langle \chi_\nu \left| \hat{L}^{R_\mu} \right| \chi_\mu \right\rangle \Bigg\}_a \tag{27}$$

This completes the outline of the GIAO–MNDO approach. The NMR chemical shift tensor is evaluated from Eqs. (18)–(21), with the first-order density matrix being determined from the CPHF equations (22)–(24) using the approximation of Eq. (27) for $H^a_{\mu\nu}$. It should be stressed that semiempirical approximations only enter the right-hand side of the CPHF equations and thus only influence the paramagnetic terms via the first-order density matrix. The present model does not explicitly account for the contributions of the core electrons to the chemical shift (due to the neglect of core electrons in MNDO-type methods), so that absolute NMR shieldings will not be reproduced. However, since core contributions to the absolute shieldings are constant to within a few ppm [56], and largely cancel when computing shifts, this shortcoming is unlikely to impede a parameterization for NMR chemical shifts, at least for first-row elements (with a 1s core).

In the implementation of the GIAO–MNDO approach, the perturbed density matrix $P^a_{\mu\nu}$, Eq. (21), is obtained by solving the CPHF equations. Unlike in the ab initio case, its determination requires only a minor fraction of the overall computational effort: relatively few one-electron integrals are needed that can be expressed in terms of overlap integrals, and the two-electron integrals are field-independent due to the MNDO approximation (see above). The derivatives $H^{ab}_{\mu\nu}$, Eq. (19), and $H^{0b}_{\mu\nu}$, Eq. (20), both contain two-center and three-center contributions. The former can be determined easily, since there are relatively few two-center terms which can be computed efficiently over Slater orbitals by any of several available techniques [59,60]. The latter represent the computational bottleneck: there are many three-center terms which are hard to compute over Slater orbitals, accounting for far more than 90% of the overall computational effort. In a semiempirical context, it is tempting to neglect these three-center contributions, but this is not possible in general, because they turn out to be important for hydrogen chemical shifts and long-range current effects. We have therefore resorted to the use of STO–4G expansions for the Slater basis functions in MNDO during the evaluation of these three-center integrals (for details see ref. [60]). This speeds up the calculation considerably, even though most of the computational effort is still spent on the three-center terms.

Calculations using standard MNDO parameters overestimate the variation of the paramagnetic contribution to the NMR chemical shifts [59] which is due to the systematic underestimation of excitation energies in MNDO (see above). This failure cannot be rectified without modifying the MNDO parameters [59]. We have therefore carried out two parameterizations for H, C, N, and O where the quantities $H^{\mathrm{MNDO}}_{\mu\nu}$ in Eq. (27) were adjusted to increase the gap between occupied and unoccupied molecular orbitals which decreases the paramagnetic contribution to the NMR chemical shifts. The first parameterization involved a total of 9 orbital

Table 5. Mean Absolute Errors of the Computed NMR Chemical Shifts (ppm)[a].

Element[b]	MA2	MA3	MB2	MB3	Range
Small set					
H (102)	0.57	0.50	0.66	0.46	31
C (93)	11.78	12.12	9.79	9.65	346
N (37)	51.13	50.51	33.53	33.70	933
O (51)	64.30	63.70	59.90	59.75	1650
Large set					
H (345)	0.95	0.66	1.60	0.73	19
C (848)	8.48	8.32	8.31	7.95	359
N (239)	41.90	41.42	28.70	28.74	977
O (168)	47.92	47.79	43.64	43.32	1698

a See text.
b Number of comparisons per element given in parentheses.

exponents and resonance parameters for H, C, N, and O which were optimized to reproduce liquid-phase chemical shift reference data from 299 small organic and inorganic molecules [59] (employing MNDO geometries and neglecting three-center terms during the parameterization for the sake of efficiency). The resulting "Method A" (MA) parameters show a significant improvement over the original MNDO parameters in chemical shift computations, a particularly good agreement with experiment being observed for ^{13}C shifts in hydrocarbons [59]. The second parameterization added the one-center one-electron energies to the previous list of parameters, for a total of 16 adjustable parameters, to allow for the tuning of both diamagnetic and paramagnetic contributions. The parameters were calibrated against experimental chemical shifts for 97 small ions and molecules (mostly gas-phase data) using B3LYP geometries and including three-center terms during the parameterization. The final "Method B" (MB) parameters turn out to be slightly superior to the MA parameters and are the recommended choice, but the quality of the computed shifts is not too different.

Table 5 lists the mean absolute errors of the calculated shifts (relative to experiment) for the small gas-phase set (97 molecules, B3LYP geometries) and the largest liquid-phase set (384 molecules, MNDO geometries). Data are given for the MA and MB parameters, both without (MA2, MB2) and with (MA3, MB3) three-center terms (for details see [60]). For both validation sets, the mean absolute errors as well as the rms errors usually remain below 5% of the total chemical shift range for each element. Closer inspection shows that a significant fraction of the error in the small set is due to small molecules with unusual bonding. Since "normal" organic molecules are represented more strongly in the larger set, the errors tend to be somewhat smaller in the latter, at least for C, N, and O. In the case of H, the experimental shifts in the large set (liquid-phase data) are partly influenced by solvation effects which cannot be completely absorbed by the parameterization; this may be the reason why the errors for H are smaller in the smaller set (gas-phase data).

Three-center terms typically contribute a few ppm to the total chemical shift for all four elements studied, but they are essential for a qualitatively correct description of hydrogen chemical shifts (and also of nucleus-independent chemical shifts, see below). These terms are less important for C, N, and O, where they can normally be omitted without significantly degrading the results. Hence, MB2 is the recommended approach for C, N, and O. MB2 calculations are fairly efficient: for example, the ^{13}C shifts of taxol (113 atoms) are computed in 117 seconds on an SGI R10000 workstation.

Nucleus-independent chemical shifts (NICS) have been introduced as an additional magnetic criterion for aromaticity [63]. They are defined as the negative magnetic shielding at some selected point in space, e.g., at a ring center. Given the large number of successful NICS studies at ab initio and DFT levels, we have applied the MNDO–MB3 approach to compute 116 NICS values for a wide range of organic molecules, including [n]annulenes, polycyclic hydrocarbons, heterocycles, cage molecules, fullerenes, and pericyclic transition states [64]. Generally we find reasonable agreement with the ab initio and DFT reference data. The semiempirical NICS values tend to be smaller in absolute value than their ab initio counterparts, but they often show similar trends. The aromatic or antiaromatic character of a given system can normally be assigned correctly on the basis of the MNDO–MB3 NICS values [64].

For a more detailed assessment of the GIAO–MNDO method, the available statistical evaluations [60,64] need to be supplemented by case studies. For example, we have confirmed that GIAO–MNDO satisfactorily reproduces the differences between the nonclassical and classical 2-norbornyl cations [56], the shifts in benzenonium and related carbocations [65], and the individual shielding tensor components in carbonyl compounds [66]. Further such validation work is in progress.

5 Analytic derivatives

Efficient explorations of potential surfaces require the derivatives of the energy with respect to the nuclear coordinates. The first derivatives (gradient) are essential for geometry optimization, while the second derivatives (harmonic force field) characterize stationary points. In principle, these derivatives can be computed either analytically or numerically. The analytic approach is generally more precise, but requires a significant coding effort. By contrast, the numerical approach is easily implemented and also trivially amenable to a coarse-grained parallelization (see below).

For variational semiempirical SCF methods (e.g., closed-shell restricted Hartree-Fock, RHF, or unrestricted Hartree-Fock, UHF) the gradient can be computed at a fraction of the cost for an SCF calculation. Analytic gradients have long been available for this case [67], but a simple finite-difference procedure with a constant density matrix and recalculated two-center integrals is also efficient. For certain nonvariational semiempirical wavefunctions (e.g., open-shell half-electron restricted Hartree-Fock, HE–RHF, or configuration interaction, CI), analytic gradients have also been introduced [68] employing a procedure for solving the CPHF equations that scales as N^4 (N basis functions). Harmonic force fields have traditionally

been determined in semiempirical programs from numerical finite differences of the corresponding gradients.

We have recently implemented analytic first and second derivatives for MNDO-type methods [59,69,70,71]. Compared with the existing codes, there is little to be gained for the gradient in variational SCF methods (see above) whereas large improvements can be achieved for the gradient in nonvariational approaches (HE–SCF, HE–CI, and small CI expansions in general). In this case, the time-determining step is the solution of the CPHF equations which can be reformulated to scale as N^3 when making use of the **Z**-vector method [72]. The chosen representation of the CPHF equations exploits the simplifications arising from the MNDO integral approximations and covers the case of fractional occupation numbers [69]. The implementation provides 18 predefined combinations of options for the solution of the CPHF equations (e.g., direct vs iterative solver, MO vs AO basis, in-core vs out-of-core treatment, alternative preconditioners) and automatically selects the best computational path for given system size and hardware configuration. As a result of these developments, dramatic speedups are observed between the previous N^4 and the new N^3 algorithms: for example, the cpu time for the gradient evaluation on an SGI Indigo R4000 workstation drops

(a) from 24049 s to 62 s in $C_{66}H_{20}$, N=284, SCF–CI time 166 s [70],

(b) from 100139 s to 222 s in $C_{96}H_{24}$, N=408, HE–SCF time 605 s [69].

It is obvious that these algorithmic improvements will greatly facilitate semiempirical studies of open-shell molecules and electronically excited states: since the evaluation of the analytic gradient is now significantly faster than the underlying SCF and CI calculations, routine full geometry optimizations become feasible for fairly large systems.

Analytic second derivatives always contain contributions from integral derivatives (direct terms) and density matrix derivatives (CPHF terms), even at the SCF level. In the ab initio case, the computational cost is often dominated by the direct terms (in typical applications to medium-size molecules) so that the analytic evaluation of the second derivatives will normally be more efficient than a numerical evaluation. By contrast, in the semiempirical case, the computational cost is always dominated by the CPHF terms since there are relatively few integral derivatives due to the MNDO approximation. While it is not trivial to derive and to code these derivatives (especially for the two-electron terms in an spd basis) [59,71], their evaluation involves little computational effort. On the other hand, the traditional finite-difference evaluation of the second derivatives in MNDO-type methods also benefits from the MNDO integral approximation and scales as MN^3 for a molecule with M atoms and N basis functions. The same scaling can be achieved in the analytic approach if the CPHF problem is formulated in the AO basis and solved iteratively [71]. It is thus the prefactors in these scaling laws (and the convergence behavior of the underlying SCF and CPHF solutions) that determine the relative speed of the two approaches. Our implementation [71] of the analytic second derivatives for MNDO-type methods turns out to be faster by factors of 4-8 compared with analogous numerical computations, and it exhibits a reliable convergence over a wider range of molecules. The asymptotic memory and disk storage requirements

can be chosen to scale as low as N^2 without significant degradation of performance. These advances clearly facilitate force constant calculations for larger molecules at the semiempirical SCF level.

6 Parallelization

Both parallel vector processors (PVP) with shared memory and massively parallel (MP) systems with distributed memory are currently used for high-performance computing (HPC). It is evident that production codes should exploit the possibilities offered by these architectures. Like other quantum-chemical software, our present semiempirical program [73] is quite large (more than 170000 lines of source code) and provides the usual core functionality as well as diverse additional options [74]. When adapting such a program to novel architectures, it is common to concentrate on the time-determining steps of standard applications first and to consider other tasks later.

Focusing on the core functionality of semiempirical codes, the most time-consuming step in an MNDO SCF–MO calculation is the solution of the secular equations, see Eq. (2). In matrix notation:

$$\mathbf{FC} = \mathbf{CE} \tag{28}$$

The solution of this eigenvalue problem is an $O(N^3)$ process. The only other $O(N^3)$ step is the calculation of the density matrix

$$\mathbf{P} = 2\ \mathbf{C}_{occ}\mathbf{C}_{occ}^T \quad , \tag{29}$$

where $\mathbf{C}_{occ}$ is an $(N \times N_{occ})$ matrix containing the N_{occ} doubly occupied closed-shell eigenvectors, and $\mathbf{C}_{occ}^T$ its transpose. Both the evaluation of the two-center integrals and the construction of the Fock matrix (see Eqs. (3)–(4)) require a computational effort that scales as $O(N^2)$. These tasks may represent a large part of the computation for smaller molecules, but the $O(N^3)$ steps are bound to dominate for large molecules and thus deserve special attention.

Traditionally the eigenvalue problem, Eq. (28), is solved by diagonalizing the Fock matrix. In the case of semiempirical SCF methods, however, this diagonalization is replaced by a pseudodiagonalization scheme whenever possible [75]. In essence, the Fock matrix is transformed from the AO basis to the MO basis (from $\mathbf{F}_{AO} = \mathbf{F}$ to $\mathbf{F}_{MO}$) using the currently available trial eigenvectors (occupied and virtual vectors being collected in $\mathbf{C}_{occ}$ and $\mathbf{C}_{virt}$, respectively):

$$\mathbf{F}_{MO} = \mathbf{C}_{occ}^T\mathbf{F}_{AO}\mathbf{C}_{virt} \quad . \tag{30}$$

The elements in the resulting $(N_{occ} \times N_{virt})$ matrix $\mathbf{F}_{MO}$ that exceed a predefined threshold are approximately annihilated by noniterative 2×2 Jacobi-type rotations (with corresponding modification of the trial eigenvectors). This pseudodiagonalization scheme is much faster than the full diagonalization and usually does not slow down SCF convergence [75]. Full diagonalizations are still required at the beginning and the end of the computation.

Measurements for medium-size molecules show that in typical applications (e.g., closed-shell RHF geometry optimizations), more than 50% of the total cpu time is

spent for matrix multiplications (see Eqs. (29)–(30)), while non-negligible portions of around 10% may be required for integral evaluation, Fock matrix construction, full diagonalization, and gradient evaluation. These are very approximate order-of-magnitude figures, of course, and actual timings will show considerable scatter depending on several factors that need not be discussed here. In any event, the share of the total cpu time spent on matrix multiplications will increase with increasing molecular size.

This is an ideal situation for PVP architectures since large matrix multiplications (DGEMM) vectorize very well and can efficiently be parallelized on shared-memory machines, running close to peak performance. Likewise, efficient library routines are available on PVP machines for full diagonalizations (EISPACK, LAPACK). The remaining minor tasks can be vectorized (Fock matrix construction) and parallelized (Fock matrix construction, integral and gradient evaluation). Hence, high performance is reached in such jobs: for example [76], a full geometry optimization of the fullerene C_{960} at the MNDO SCF level runs at a speed of 23.9 GFLOPS/s on the NEC SX-4/16 at CSCS (75% of the hardware limit of 32 GLOPS/s for 16 cpus, speedup factor of 13.8 relative to 1 cpu, wallclock time 2319 s).

The situation is less advantageous for distributed-memory MP systems. A coarse-grained parallelization is possible for certain applications where essentially independent jobs can be distributed over different nodes. Examples of this kind include the finite-difference evaluation of gradients and force constants as well as reaction path calculations. These applications show almost linear speedup with the number of processors [74], but they are not prevalent in practice, and the size of molecules that can be handled in this manner is limited by the memory available at each single node [74].

It is thus necessary to attempt a fine-grained parallelization of the basic MNDO SCF procedure on MP systems such that the relevant data are distributed over all p available processors. We start by considering the square symmetric Fock matrix $\mathbf{F}$ which is distributed by rows over the processors. The division of $\mathbf{F}$ is determined once and for all at the beginning of the computation such that the boundaries are always between basis functions of different atoms. Hence, each processor conceptually owns a partial square Fock matrix $\mathbf{F}_p$ of dimension $(N_p \times N)$ with $N_p \approx N/p$. This partial square matrix is associated with a subset $\{A_p\}$ of atom pairs, i.e. those involving atoms whose basis functions correspond to the row indices of $\mathbf{F}_p$. Inspection of Eqs. (3)–(4) shows that the construction of $\mathbf{F}_p$ requires only the one-electron integrals from the corresponding partial square matrix $\mathbf{H}_p$ and the two-electron integrals $\mathbf{G}_p$ from the subset $\{A_p\}$ of atom pairs. Moreover, the calculation of $\mathbf{F}_p$ from Eqs. (3)–(4) employs only two-center density matrix elements that belong to the corresponding partial square matrix $\mathbf{P}_p$ whereas the required one-center density matrix elements $\mathbf{P}^{AA}$ refer to all atoms in the molecule.

Based on these observations, the partial Fock matrix $\mathbf{F}_p$ can be constructed locally at the corresponding processor using one-electron and two-electron integrals which can also be evaluated locally in an integral-direct manner. The one-center density matrix elements $\mathbf{P}^{AA}$ must be communicated globally, whereas the partial density matrix $\mathbf{P}_p$ needs to be communicated only to the associated processor. This

scheme [74] of constructing the Fock matrix in parallel exploits the simplifications inherent to the MNDO approximation.

The remaining and most time-consuming steps of the MNDO SCF procedure, Eqs. (28)–(30), are handled by calls to standard parallel library routines for matrix multiplication (PBLAS) and full diagonalization (ScaLAPACK), with a corresponding adaptation of the data structures [77]. Benchmark calculations with the resulting parallel code have been carried out some time ago on a Cray T3D system using up to 64 nodes [77]. These calculations showed that large molecules can be treated on this modest machine: Using 64 nodes, the fullerene C_{960} required less than 44 MB per node, with wallclock times of 2275 s and 13 s for the MNDO SCF energy and gradient evaluation, respectively. The MNDO geometry optimization of the fullerene C_{540} showed a speedup of 1.59 when going from 16 to 32 nodes, and another speedup of 1.35 upon moving to 64 nodes; these factors are mostly determined by the performance of the library routine (PSGEMM and PSSYEVX) where more than 75% of the cpu time are spent in this case. Apparently, the fullerene examples chosen are two small (matrix dimensions of 2160 for C_{540} and 3840 for C_{960}) to obtain a convincing scaling behavior from the linear algebra routines.

In summary, our semiempirical code has been ported both to shared-memory PVP and distributed-memory MP platforms. Both versions have been tested for molecules containing up to about 1000 non-hydrogen atoms. The performance is excellent on PVP machines up to 16 nodes, and satisfactory on MP systems up to 64 nodes. In the latter case, the code would seem to require further optimization, and larger molecules would need to be tested with more than 64 nodes. On the other hand, it is probably not worthwhile to do conventional semiempirical SCF calculations (with diagonalization) on molecules with more than 1000 atoms, even on MP systems, because alternative theoretical and algorithmic approaches are available for such large systems. This will be discussed in the next section.

7 Linear scaling and combined QM/MM approaches

In recent years, it has been a common goal of many groups to perform theoretical calculations on ever larger molecules. In the field of quantum chemistry, one common theme in this endeavour has been to exploit the locality of interactions whenever possible and to introduce numerically well-controlled simplifications for long-range interactions. This leads to computational procedures where the scaling of the computational effort with system size is improved over the formal scaling that characterizes a given method. The ultimate goal of such research are methods that exhibit linear scaling with system size.

It is clearly beyond the scope of this article to survey the widespread activities on linear scaling approaches. Some of the ab initio and DFT work has been reviewed recently [78]. In semiempirical quantum chemistry, the primary objective of the linear scaling algorithms is to avoid the diagonalization bottleneck, i.e. to avoid the steps that scale as N^3 (see above). Three different approaches have been proposed which accomplish this goal:

(a) Localized molecular orbital (LMO) techniques [79],

(b) Divide-and-conquer methods [80,81,82],

(c) Conjugate gradient density matrix search (CG–DMS) [83,84,85].

In the LMO approach, 2 × 2 rotations are applied to annihilate the interactions between occupied and virtual LMOs that are located within a certain cutoff radius (typically 8–10 Å), whereas all other interactions are considered to be negligible and therefore not treated. Small numerical errors result from these approximations, but they can be controlled by a renormalization of the LMOs and a suitable choice of the cutoff radius. The SCF procedure for LMOs is found to converge even faster than the conventional one. Using this algorithm, it is possible to carry out semiempirical SCF–MO calculations on standard workstations for proteins with several thousand atoms. The largest system treated in the original paper is rhizomucor miehei lipase with 4037 atoms [79].

The divide-and-conquer methods are based on a partitioning of the density matrix. The overall electronic structure calculation is decomposed into a series of relatively inexpensive calculations for a set of smaller, overlapping subsystems, each of which involves the diagonalization of a comparatively small Fock matrix. A global description of the full system is then obtained by combining the information from all subsystem density matrices. The accuracy of this approach can be controlled by the manner in which the system is partitioned. Proteins that have been treated by the divide-and-conquer method include HIV protease 6-mer with 9378 atoms [81].

The CG–DMS method avoids diagonalizations by using a direct conjugate gradient search for the density matrix. An acceptable density matrix must be normalized and idempotent, and it must commute with the Fock matrix after SCF convergence. In the CG–DMS treatment, a suitable functional of the density matrix is minimized with respect to the density matrix such that the resulting density matrix satisfies the above criteria; idempotency is enforced through McWeeny purification transformations. By neglecting density and Fock matrix elements close to zero, sparse matrices are obtained, and linear scaling can be approached by applying sparse matrix techniques. The accuracy of this algorithm depends on the chosen cutoffs. Initial benchmarks with this method include nucleic acids up to 6304 atoms and polyglycine chains up to 19995 atoms [85].

The methods outlined above share several common features. First, all of them introduce some approximations so that the results from such treatments will show some deviations from the conventional results obtained by a full diagonalization. These deviations will increase when cutoffs are made less stringent to speed up the calculations, and it is obviously necessary to ensure that they remain tolerable. Secondly, all three methods require some overhead so that the conventional calculations with full diagonalization remain faster for small molecules. The crossover point beyond which the described algorithms become faster depends on a number of factors (e.g., the chosen cutoffs and the molecular shapes), but 200–300 atoms seems to be a typical range. In view of this situation, it makes little sense to invest in the optimization of conventional codes to treat systems with more than thousand atoms more efficiently (see above). Finally, it should be kept in mind that none of these algorithms can exhibit true linear scaling without addressing other parts of semiempirical computations that scale formally as N^2, such as integral evaluation.

However, in practice this is only a minor point since the N^2 steps are generally very fast so that the removal of the diagonalization bottleneck is indeed of decisive importance.

The strength of the semiempirical "linear scaling" methods is most obvious for large systems with long-range charge transfer or long-range charge fluctuations, which may, for example, occur in proteins. Such effects can be captured by these approaches. On the other hand, there are many processes where the electronically active part of a large system is more localized, as in many chemical reactions or in localized electronic excitations. For such systems, an alternative theoretical treatment may be appropriate where the active center is described by quantum mechanics (QM) and the environment by molecular mechanics (MM). Such combined QM/MM methods are not the subject of this article, but they are briefly mentioned here to put the linear scaling developments into perspective.

Combined QM/MM methods (for reviews see refs. [86,87,88,89]) are computationally much less demanding than pure QM methods, even when compared with semiempirical linear scaling approaches. In addition, they offer the possibility for a system-specific modeling of large systems, by selecting suitable QM and MM components for the chemical problem being studied. On the other hand, QM/MM methods require a careful definition of the QM/MM boundary (especially when covalent bonds need to be cut), a physically sensible QM/MM coupling scheme, and possibly a calibration of QM/MM interaction terms [86,87,88,89,90,91]. Moreover, the standard QM/MM approaches do not allow for charge transfer between the QM and MM regions.

At present, it seems that the linear scaling QM methods and the combined QM/MM methods are complementary since they may be used to address different questions. Therefore, our group is active in both these areas which are expected to coexist and supplement each other for some time to come.

Due to the advances in linear scaling and QM/MM methods, the semiempirical methods of quantum chemistry have come to the point where they can be applied to complex systems with thousands of atoms such as enzymes. A relatively large number of enzymatic reactions has already been studied at the semiempirical QM/MM level (for a survey see ref. [88]), and analogous investigations with semiempirical linear scaling approaches are anticipated.

8 Conclusions

The methodological and algorithmic developments outlined in this article promise to open new areas of application for semiempirical methods which will therefore continue to be valuable tools for studying electronic effects in large molecules. Whenever technically feasible, such investigations should be supplemented with appropriate higher-level calculations because the synergetic use of several computational tools is often expected to provide the best computational solution for a given chemical problem.

Acknowledgements

The author wishes to thank his colleagues and coworkers whose names are listed in the references for their essential contributions to the work described. Special thanks are due to F. Terstegen for his help with the manuscript.

References

1. M.J.S. Dewar, *The Molecular Orbital Theory of Organic Chemistry*, McGraw-Hill, New York, (1969).
2. J.A. Pople and D.L. Beveridge,*Approximate Molecular Orbital Theory*, Academic Press, New York, (1970).
3. J.N. Murrell and A.J. Harget, *Semiempirical Self-Consistent-Field Molecular Orbital Theory of Molecules*, Wiley, New York, (1972).
4. G.A. Segal, Ed., *Modern Theoretical Chemistry*, Plenum, New York, (1977), Vols. 7-8.
5. M. Scholz and H.J. Köhler,*Quantenchemie*, Hüthig, Heidelberg, (1981), Vol. 3.
6. T. Clark,*A Handbook of Computational Chemistry*, Wiley, New York, (1985).
7. M.J.S. Dewar, Science **187**, 1037-1044 (1975).
8. K. Jug, Theor. Chim. Acta **54**, 263-300 (1980).
9. M.J.S. Dewar, J. Phys. Chem. **89**, 2145-2150 (1985).
10. W. Thiel, Tetrahedron **44**, 7393-7408 (1988).
11. J.J.P. Stewart, J. Comp.-Aided Mol. Design **4**, 1-103 (1990).
12. J.J.P. Stewart, *in Reviews in Computational Chemistry* (Eds. K.B. Lipkowitz and D.B. Boyd), Vol. **1**, VCH Publishers, New York, (1990), pp. 45-81.
13. M.C. Zerner, *in Reviews in Computational Chemistry* (Eds. K.B. Lipkowitz and D.B. Boyd), Vol. **2**, VCH Publishers, New York, (1991), pp. 313-365.
14. W. Thiel, Adv. Chem. Phys. **93**, 703-757 (1996).
15. J.A. Pople, D.P. Santry, and G.A. Segal, J. Chem. Phys. **43**, S129-S135 (1965).
16. J.A. Pople and G.A. Segal, J. Chem. Phys. **44**, 3289-3296 (1966).
17. J. Del Bene and H.H. Jaffé, J. Chem. Phys. **48**, 1807-1813 (1968).
18. J.A. Pople, D.L. Beveridge, and P.A. Dobosh, J. Chem. Phys. **47**, 2026-2033 (1967).
19. R.C. Bingham, M.J.S. Dewar, and D.H. Lo, J. Am. Chem. Soc. **97**, 1285-1293 (1975).
20. M.C. Zerner and J. Ridley, Theor. Chim. Acta **32**, 111-134 (1973).
21. A.D. Bacon and M.C. Zerner, Theor. Chim. Acta **53**, 21-54 (1979).
22. D.N. Nanda and K. Jug, Theor. Chim. Acta **57**, 95-106 (1980).
23. K. Jug, R. Iffert, and J. Schulz, Int. J. Quantum Chem. **32**, 265-277 (1987).
24. B. Ahlswede and K. Jug, J. Comput. Chem. **20**, 563-571 (1999).
25. B. Ahlswede and K. Jug, J. Comput. Chem. **20**, 572-578 (1999).
26. M.J.S. Dewar and W. Thiel, J. Am. Chem. Soc. **99**, 4899-4907 (1977).
27. M.J.S. Dewar and W. Thiel, J. Am. Chem. Soc. **99**, 4907-4917 (1977).
28. W. Thiel, J. Am. Chem. Soc. **103**, 1413-1420 (1981).
29. M.J.S. Dewar, E. Zoebisch, E.F. Healy, and J.J.P. Stewart, J. Am. Chem. Soc.

107, 3902-3909 (1985).
30. J.J.P. Stewart, J. Comput. Chem. **10**, 209-220 (1989); J. Comput. Chem. **10**, 221-264 (1989).
31. M.J.S. Dewar, C. Jie, and J. Yu, Tetrahedron **49**, 5003-5038 (1993).
32. A.J. Holder, R.D. Dennington, and C. Jie, Tetrahedron **50**, 627-638 (1994).
33. W. Thiel and A.A. Voityuk, Theor. Chim. Acta **81**, 391-404 (1992); Theor. Chim. Acta **93**, 315 (1996).
34. W. Thiel and A.A. Voityuk, J. Phys. Chem. **100**, 616-626 (1996).
35. SPARTAN 4.0 (1995), Wavefunction Inc., 18401 Von Karman Avenue, Irvine, CA 92715.
36. A.A. Voityuk, M.C. Zerner, and N. Rösch, J. Phys. Chem. **A 103**, 4553-4559 (1999).
37. M.J.S. Dewar and W. Thiel, Theor. Chim. Acta **46**, 89-104 (1977).
38. L.A. Curtiss, K. Raghavachari, P.C. Redfern, and J.A. Pople, J. Chem. Phys. **106**, 1063-1079 (1997).
39. W. Thiel, *in "Computational Thermochemistry"* (Eds. K.K. Irikura and D.J. Frurip), ACS Symposium Series 667, American Chemical Society, Washington, DC, (1998), pp.142-161.
40. C. Kollmar and M.C. Böhm, Theor. Chim. Acta **92**, 13-47 (1995).
41. S. de Bruijn, Stud. Phys. Theor. Chem. **21**, 251-272 (1982).
42. A.R. Gregory and M.N. Paddon-Row, J. Am. Chem. Soc. **98**, 7521-7523 (1976).
43. M.V. Basilevsky, A.G. Shamov, and V.A. Tikhomirov, J. Am. Chem. Soc. **99**, 1369-1372 (1977).
44. P. Caramella, K.N. Houk, and L.N. Domelsmith, J. Am. Chem. Soc. **99**, 4511-4514 (1977).
45. J. Spanget-Larsen, Theor. Chim. Acta **98**, 137-153 (1997).
46. W. Weber, Ph.D. Thesis, Universität Zürich, (1996).
47. W. Weber and W. Thiel, Theor. Chem. Acc. , in press (2000).
48. J. Spanget-Larsen, Theor. Chim. Acta **55**, 165-172 (1980).
49. C. Kollmar, Chem. Phys. Lett. **269**, 215-221 (1997).
50. M.J. Filatov, O.V. Gritsenko, and G.M. Zhidomirov, Theor. Chim. Acta **72**, 211-222 (1987).
51. M. Kolb, Ph.D. Thesis, Universität Wuppertal, (1991).
52. M. Kolb and W. Thiel, J. Comput. Chem. **14**, 775-789 (1993).
53. M. Beck, K. Möhle, M. Paech, R. Steiger, W. Thiel, E.U. Wallenborn, and W. Weber, to be published.
54. R. Ditchfield, Mol. Phys. **27**, 789-807 (1974).
55. W. Kutzelnigg, Isr. J. Chem. **19**, 193-200 (1980).
56. W. Kutzelnigg, U. Fleischer, and M. Schindler, NMR Basic Principles Progress **23**, 165-262 (1990).
57. J. Gauss, Ber. Bunsenges. Phys. Chem. **99**, 1001-1008 (1995).
58. G. Schreckenbach and T. Ziegler, Theor. Chem. Acc. **99**, 71-82 (1998).
59. S. Patchkovskii, Ph.D. Thesis, Universität Zürich, (1997).
60. S. Patchkovskii and W. Thiel, J. Comput. Chem. **20**, 1220-1245 (1999).
61. K. Wolinski, J.F. Hinton, and P. Pulay, J. Am. Chem. Soc. **112**, 8251-8260

(1990).
62. F. London, J. Phys. Radium **8**, 397 (1937).
63. P.v.R. Schleyer, C. Maerker, A. Dransfeld, H. Jiao, and N.J.R. van Eikema Hommes, J. Am. Chem. Soc. **118**, 6317-6318 (1996).
64. S. Patchovskii and W. Thiel, J. Mol. Model. , in press (2000).
65. M. Kollwitz and J. Gauss, Chem. Phys. Lett. **260**, 639-646 (1996).
66. N. Gonzales and J. Simons, Int. J. Quantum Chem. **63**, 875-894 (1997).
67. M.J.S. Dewar and Y. Yamaguchi, Comput. Chem. **2**, 25-29 (1978).
68. M.J.S. Dewar and D.A. Liotard, J. Mol. Struct. (Theochem) **206**, 123-133 (1990).
69. S. Patchkovskii and W. Thiel, Theoret. Chim. Acta **93**, 87-99 (1996).
70. S. Patchkovskii and W. Thiel, Theor. Chem. Acc. **98**, 1-4 (1997).
71. S. Patchkovskii and W. Thiel, J. Comput. Chem. **17**, 1318-1327 (1996).
72. N.C. Handy and H.F. Schaefer, J. Chem. Phys. **81**, 5031-5033 (1984).
73. W. Thiel, Program MNDO99, Mülheim, (1999).
74. W. Thiel and D.G. Green, *in: Methods and Techniques in Computational Chemistry, METECC-95* (Eds. E. Clementi and G. Corongiu), STEF, Cagliari, (1995), pp. 141-168.
75. J.J.P. Stewart, P. Csaszar, and P. Pulay, J. Comput. Chem. **3**, 227-228 (1982).
76. C. Henriet, NEC Corporation, private communication (September 1998).
77. R. Graham, Cray Research, private communication (November 1995).
78. G.E. Scuseria, J. Phys. Chem. **A 103**, 4782-4790 (1999).
79. J.J.P. Stewart, Int. J. Quantum Chem. **58**, 133-146 (1996).
80. W. Yang and T.-S. Lee, J. Chem. Phys. **103**, 5674-5678 (1995).
81. T.S. Lee, D.M. York, and W. Yang, J. Chem. Phys. **105**, 2744-2750 (1996).
82. S.L. Dixon and K.M. Merz, J. Chem. Phys. **104**, 6643-6649 (1996).
83. X.-P. Li, R.W. Nunes, and D. Vanderbilt, Phys. Rev. **B 47**, 10891-10894 (1993).
84. J.M. Millam and G.E. Scuseria, J. Chem. Phys. **106**, 5569-5577 (1997).
85. A.D. Daniels, J.M. Millam and G.E. Scuseria, J. Chem. Phys. **107**, 425-431 (1997).
86. J. Aqvist and A. Warshel, Chem. Rev. **93**, 2523-2544 (1993).
87. J. Gao, *in Reviews in Computational Chemistry* (Eds. K.B. Lipkowitz and D.B. Boyd), Vol. 7, VCH Publishers, New York, (1995), pp. 119-185.
88. T.Z. Mordasini and W. Thiel, Chimia **52**, 288-291 (1998).
89. G. Monard and K.M. Merz, Acc. Chem. Res. **32**, 904-911 (1999).
90. D. Bakowies and W. Thiel, J. Phys. Chem. **100**, 10580-10594 (1996).
91. I. Antes and W. Thiel, J. Phys. Chem. **A 103**, 9290-9295 (1999).